Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005

A Study of Spatial Cognition in an Immersive Virtual Audio Environment: Comparing Blind and Blindfolded Individuals

Amandine Afonso, Brian FG Katz, Alan Blum, Christian Jacquemin, and Michel Denis

LIMSI-CNRS
BP 133
F-91403 Orsay
France
amandine.afonso@limsi.fr katz@limsi.fr

ABSTRACT

This study presents the combined efforts of three research groups toward the investigation of a cognitive issue through the development and implementation of a general purpose VR environment that incorporates a high quality virtual 3D audio interface. The psychological aspects of the study concern mechanisms involved in spatial cognition, in particular to determine how a verbal description of an environment or the active exploration of that environment affects the building of a mental spatial representation. Another point is to investigate the role of vision by observing whether or not participants without vision (blind from birth, late blind or blindfolded sighted individuals) can benefit from these two learning modalities. This paper presents the preliminary results of this study. Additionally is a description of the generic toolkit and companion architecture that has been developed and used for modeling the environment and interface in a cohesive manner. Details for generating an immersive multimodal experimental environment for this platform are also included.

1. INTRODUCTION

In recent years, virtual reality (VR) techniques have developed considerably in the domain of computer sciences. These techniques allow users to interact in as natural as possible a manner with data made available to their sensory experience, visual in the majority of the applications, but also auditory and kinesthetic in immerging research contexts [1][2]. There are great benefits for studies of human behavior with the inclusion of this resource in current paradigms. The special value of VR environments is to allow the investigation of human behavior with people immersed in realistic controlled interactive contexts, without the physical constraints and costs of building such real contexts. VR is an invaluable tool for creating situations that make the study of human behavior easier by expanding the scope of experimental research [3][4].

The work reported here aims at illustrating the capacity of VR as a research tool for the analysis of human cognition and behavior in complex environments. The use of an audio VR platform was of special relevance with respect to the study's purpose of exposing human participants to auditory scenes containing complex sets of spatially organized data, and to allow interaction with elements of the scenes. The immersive character of the VR experience gives participants the sense that the auditory objects they perceive are present in the room and that despite their movements they are within a stable and consistent spatial environment. Through this perceptive stability, and the flexibility of dynamic scene interactions, the VR environment can greatly aid studies in the domain of cognitive/behavioral sciences, in particular in the study of the loop connecting perception, cognition, and action [5].

The significant feature of the present research was to undertake an empirical investigation of a cognitive issue by implementing a general purpose VR environment which incorporated a high quality virtual 3D audio interface. In the following sections, the psychological context shall be discussed, following which the generic experimental platform developed for multimodal interactions will be presented in detail. The remaining sections of the paper report the experimental procedure and the results obtained.

2. VISUAL IMAGERY AND SPATIAL COGNITION

Visual imagery can be defined as the representation of perceptual information in the absence of visual input [6]. In order to assess whether visual experience is a pre-requisite for image formation, many studies have focused on the analysis of visual imagery in congenitally blind participants. However, only few studies have described how visual experience affects the metric properties of the mental representations of space.

The experiment presented here is intended to further investigate the effect of prior visual experience on mental imagery as it pertains to spatial representations. The first study addressing this issue was reported by Kerr who, using haptic learning of a spatial configuration and a mental scanning task, found a strong relation between scanning times and the distance to travel mentally for sighted people as well as for blind people, whether they were blind from birth or became blind later in their life [7]. Nevertheless, the time needed to scan different distances for blind persons was significantly longer than for blindfolded persons. In contrast, Röder and Rösler found that the chronometric performance of blind and blindfolded sighted people did not differ in a mental scanning paradigm involving haptic learning of the configuration [8].

A previous study has been carried out by several of the present authors to investigate mental scanning, as well as performance in distance comparison tasks, following either verbal or haptic learning, as applied to blind (congenital or late) and sighted (blindfolded or with unobstructed vision) individuals [9][Afonso *et al*, in preparation]. The main conclusion was that blind individuals, like sighted, were able to generate an accurate mental representation of an environment which they acquired either through a verbal description or through haptic exploration of a small-scale spatial configuration. However, blind people required significantly more time.

The working hypothesis is that the mental representation of spatial configurations may result from different strategies. A sighted person may use mental representations that are iconic in

nature, whereas blind individuals could better memorize sensorimotor contingencies. De Beni and Cornoldi have suggested that visual images, even in the sighted, may be representations based on information collected through a range of different sensory modalities [10].

Historically, studies on mental imagery have shown that the main property of mental images is their structural isomorphism to the configurations from which they are generated [11][12]. More particularly, mental scanning studies conducted by Kosslyn, Ball, and Reiser [13], and later adapted by Denis and Cocude [14], have shown that mental representations constructed from visual experiences (or from verbal descriptions) are analog reflections of corresponding configurations. Analysis of chronometric data has shown that the greater the distance separating two points of a configuration, the longer the corresponding mental scanning time. This finding has been considered as support for the hypothesis that visual images genuinely reflect the metric properties of objects, and that they undergo the same functional constraints as those which apply to perception (for a review see [15]). This fact has been used as an indication of the visual character of spatial representations constructed from visual perception as well as those constructed from verbal descriptions.

3. EXPERIMENTAL CONCEPT

A large-scale immersive audio virtual environment (no visual feed-back, sighted participants are blindfolded) was created in which participants could explore and interact with virtual sound objects located within a room. The objective was to assess the effect of visual experience on mental spatial imagery.

Two learning conditions were contrasted, one expected to induce a mental representation of an iconic nature, and the other generating a mental representation based on perception/action couplings. The comparison was intended to help identify which learning modality generates the most accurate representation of a spatial configuration, as well as to collect information on the perception of a world by blind and sighted people interacting with a three-dimensional audio environment.

The task performed by the participants was the manual recreation of a known sound scene. Through this experiment, several questions regarding mental imagery were investigated which could not have been possible without this technique. This paper reports the preliminary results from this experiment. Additional tasks were also included, involving the proposition of different virtual scenes exhibiting small and large metric changes followed by questions and the ability to correct the scene. Finally, mental scanning tests were conducted.

4. EXPERIMENTAL PLATFORM

Previous experiments using 3D audio for sound localization experiments or for building audio interfaces for blind people rely on platforms dedicated to spatialized audio rendering that have little or no graphic output. Whether with real sound sources [16] or within a virtual environment [17], user tracking is necessary and a minimal geometrical model of the scene must be updated in real-time. Even though the same base components exist here (tracking and scene representation), our approach to virtual audio modeling is different because it relies on a tool with full capabilities for multimedia 3D effects, behavioral modeling, and interaction: Virtual Choreographer (*VirChor*) [18]. The reason for this choice is due to the complexity of our experimental setup that requires the experimenter to monitor accurately the location of the participant and active audio sources. In addition, the complexity of the protocol and its progressive definition through real experiments called for an open scripting language that could be easily modified on-site. In this section, we present both the architectural and software designs for rendering spatialized sound and graphics as well as controlling the interface behavior in response to participant and experimenter inputs.

4.1. Audio in Virtual / Augmented Reality

The fundamental requirements of the experiment are an accurate representation of a sonic scene with which the participant can navigate and interact. The majority of audio components in virtual/augmented reality environments are relatively limited in their quality and resolution. Many systems still implement only stereo panning of sound sources. Most current implementations of 3D graphical rendering used in games and virtual or augmented environments for collaborative work would benefit from a richer sonic rendering.

While the context of the study is based around a purely auditory environment, there is an inherent geometry associated. The various sound sources can only be correctly spatialized in a geometrical framework. The physical space must be represented. The positions of the participant and sound sources must be constantly updated within the 3D geometry in order to maintain the correct relative locations of the sound sources with respect to the participant within the environment.

For real-time experimental control the experimenter is aided by a visual feedback of the entire scene that mirrors the current status of the internal representation: active sound sources and their locations, location of the participant in the virtual scene.

4.2. System Overview

The multimedia scene in which the experiment takes place consists of a room (both physical and virtual) in which six virtual sound objects are located. The participant is equipped with a head-tracker device, mounted on a pair of stereophonic headphones, as well as a handheld tracked pointing device. The experimenter controls the course of the experiment and can constantly verify the status of the system on a computer display, an example of which is shown in Figure 1. The left panel shows the current subjective view of the participant and the right panel of the figure presents an overview of the room. The scene consists primarily of the six sound sources (represented by numbered spheres, where red spheres indicate that the sound source is active), the participant (head), and the pointing device (arrow). The reference scene consists of the spheres located on a circle, also visible in the display. Even though the participant only experiences the auditory component of the model, the actual experimental room has been modeled (and photo texture mapped). This allows the experimenter to better interpret the participants placement and orientation in the scene. In addition, collision detection is used to warn the participants (through an auditory alert) if they approach the boundaries of the physical room or the limits of the tracking system.

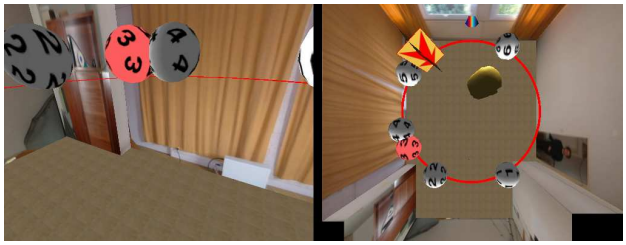

Figure 1. *Screenshot of the combined subjective and overview displays).*

The general flow of information within the system is shown in Figure 2. The six degree-of-freedom (6 DOF) tracking system is polled by *Max/MSP* [19] for the current position of the participant. The positional information is then passed to the modeler, VirChor. After integrating the external positional updates, experimenter controls, and internal interactions, VirChor sends updated relative source positions (spherical coordinates in the participant's reference frame according to the subjective view in Figure 1) and audio controls to Max/MSP. These parameters are then used to control the audio rendering. The spatialized audio is finally delivered to the participant via headphones.

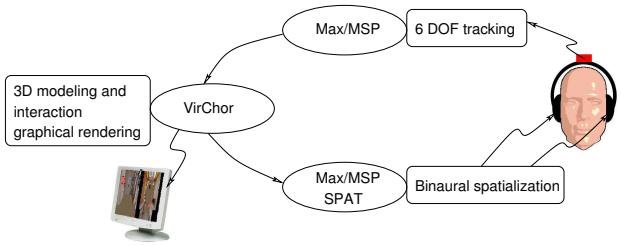

Figure 2. Overview of the architecture.

4.3. Scene Model

The key issue in designing the experiment was to provide the participant and the experimenter with a reactive interface that would implement a scenario of multiple stages with tasks of orientation and localization. In addition, the control of the experiment had to be restrained to a set of minimal operations in order to avoid burdening the experimenter with complex control procedures.

The experimental setup was installed in an existing room, as shown in Figure 3. The experimental room was approximately 4x6 m of which the majority was accessible by the participants. A MIDI interface, used by the experimenter to alter the current state of the experiment (via changes to the internal states of the spheres) is indicated on the figure, as well as the location of the visual feedback screen and machine room outside the experimental room. The configuration of the reference virtual sound scene, central reference point, and physical reference point (chair) used during the experiment are also shown.

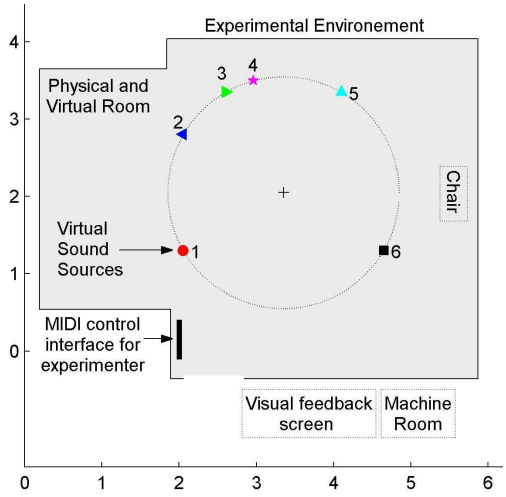

Figure 3. *Physical and virtual scene of the experiment (shown on a meter scale).*

The scene was made of three types of objects that all belonged to the class of Geometrical objects in *VirChor*: physical components (walls and floor), collision detection devices (used to create alerts to avoid wall contact), and sound components (sonic spheres). The role of collision detection devices is to respond to a sensor entering their bounding volume and to emit a trigger, here initiating an audio alert. Most of the behavioral capabilities of the scene are located on the sonic spheres. The behavior of the spheres is regulated via their internal states (controlled by the experimenter) and consists of scripts that trigger sound outputs, cascaded sphere activations, random sphere displacements, or user-controlled sphere positioning. Figure 4 shows an example of the definition of a sonic sphere: a geometrical textured sphere that carries sound properties. Part of the associated scripting is also provided, showing the cascading of messages based on internal states.

```
<node id="sphere 1">
  <script id="script sphere 1">
    <!---- Experiment 0a: each sphere is active in turn ---->
    <command>
      <trigger type="message_event" value="click" state="Exp0a"
          bool_operator="==" />
       <action>
         <send_message value="start sphere 1" />
           <target type="single_node" value="#main sphere" />
         </action>
       <action>
         <set_internal_state value="Exp0a_ongoing" />
          <target type="single_node" value="#sphere 1" />
        </action>
       <action> <write_console value="Msg: Starting Source1" /> </action>
    </command>
  </script>
  <sound id="snd sphere 1" xlink:href="{#file1}" type="soundloop"
   fade_distance="0.0" fade_power="0" level="{#deflev1}" source="1"
   begin="10000" end="100000" period="10.0" dur="10.0">
  </sound>
  <sphere id="geo sphere 1" radius="0.25" segments=20>
    <texture encoding="jpeg" env_mode="modulate" tile_s="1" tile_t="1"
   xlink:href="textures/1.jpg" id="1"/>
  </sphere>
</node>
```

Figure 4. *Example VirChor scene script for an audio/graphic object node.*

4.4. Behavior and 3D Audio/Graphic Rendering

VirChor uses an XML syntax for scene modeling which is very similar to the graphic component of X3D [20]. The *Contigra* project proposes an extension of X3D to include audio [21] and behavior [22]. In Contigra, audio scenes are described independently from the graphical scene. This is so that the same audio scene can be integrated into several graphic scenes.

The scene graph structure within VirChor is based on the concept of a unique and cohesive hierarchy framework of scene nodes [23]. Nodes can be comprised of properties directed toward rendering (graphical, auditory, etc.) and behavioral scripts.

Behavior within VirChor is modeled through internal message exchange between scene nodes or external communication between these elements and networked applications via UDP. For example, a distributed architecture, employing UDP inter-communications, allows the graphical rendering and audio rendering to be performed on separate machines. Message reception by scene nodes can be controlled by internal node states, triggers, cascaded message transmissions, or scene node modification. These messages can be real-time (interactive with the experimenter or participant), scheduled, or mixed as with a launched series of scheduled events. Scene control is performed through partial XML elements which define the parameter updates. The syntax for internal and external communications is identical and straightforwardly derived from XML element syntax. Figure 5 illustrates the information flow between the components of VirChor.

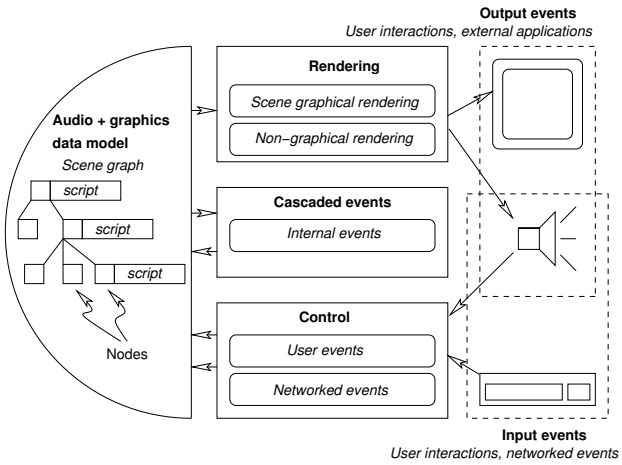

Figure 5. *Architecture for graphical and sonic rendering.*

All control messages (internal or external from other applications and from peripherals) are time-stamp logged in order to allow for post-treatment analyses, motion graph plotting, and scene replay. During replay, all of the sound scene is automatically reproduced as all external control events, including the tracking system, are replayed into VirChor which then reacts as during the actual experiment. Time dilation is also possible, allowing variations in replay speed.

4.5. Latency Measurement

Measurement of the total system latency (TSL), as defined by Wenzel [24], was made. The objective was to estimate the time between a movement in the real world and its effective consequence on the sound scene diffused through the headphones of the participant. Within the system architecture, positional information is obtained by periodically (every 70 ms) polling the tracker hardware. This request is initiated within Max/MSP via the serial communication port (9.6 kBaud). Each polling cycle begins with a message prompting the tracker for the current position, which is then read from the serial buffer 10 ms later. In general, it was observed that this buffer gives the previous position so that the actual current position is updated in the next cycle, resulting in a delay of 80 ms. A communication message bounce, via UDP, was performed between two VirChor modules on a single machine, producing a latency of 28 ms (equivalent to twice the current frame refresh rate of 70 Hz).

After determining several individual latencies, the next step consists in estimating the complete TSL. The participant's tracker was placed at a distance far from a single active source, playing a 16 kHz pure tone. The tracking-polling was then paused. The tracker was then placed very near the virtual source position, so that when the position was updated the rendered audio level would increase dramatically. The tracker-polling was then re-initiated, t_0. Using this method, a simple threshold detector on the audio output, indicating that the scene has been updated, can be used to terminate the measurement cycle (t = end time). Ten measurement repetitions resulted in a mean total latency ($t - t_0$) of 248 ms (standard deviation, $\sigma = 28$ ms). To be more precise, one should correct for the time due to the threshold analysis of the sound level, which required the passage of the signal through an analog to digital converter. The system was configured to use a 512 sample input buffer and a 64 sample analyses buffer, resulting in a 13 ms delay at the sample rate of 44.1 kHz.

The final TSL estimation is 235 ms. This is a non-negligible value. The perceptual effects of this latency are discussed in the Subjective Questionnaire Results section.

4.6. Sound Processing Architecture

Sound spatialization was performed using the Max/MSP environment and IRCAM's reverberation and spatialization library, *Spat*. A passive interface was developed which allows for all audio rendering to be controlled by external communications with VirChor. It is important to note that the rendering method used, binaural synthesis (described in section 4.7), is computationally intensive, increasing with the number of sources. To reduce computational load while maintaining scene flexibility, a hierarchical audio scene structure was created which includes provisions for multi-user cooperative environments. The audio scene tree comprises three levels: room, user, and source. This concept makes use of Spat's "shared reverberation" calculation which individually renders the direct sound and early reflection but creates a single reverberant tail. Therefore the calculation of the late reverberation part, which is considered homogeneous, can be done only once for a monaural mix of all active sources signals. Using this scheme, all sources within a given "room" acoustic use a shared reverb. If multiple room acoustics are desired for other users, additional "rooms" must be defined.

The balance between direct and reverberant sound energy is useful in the perception of source distance [25]. It has also been observed that the reverberant energy, and especially a diffuse reverberant field, can negatively affect source localization. As this study was primarily concerned with a spatially precise rendering, rather than a realistic room acoustic experience, the reverberant energy was somewhat limited. Omitting the room effect creates an "anechoic" environment, which is not habitual for most people. It was decided in this study to create a more realistic environment for which the room effect was included. A room effect, characterized by a reverberation time of 2 s, was employed. To counteract the negative effect on source localization, the direct to reverberant ratio was defined as 10 dB at 1 m.

4.7. Binaural Synthesis

Binaural synthesis is an audio presentation technique that attempts to present spatially encoded audio directly at the ear canal of the user. Natural spatial encoding is performed by the natural filtering of sound arriving at the ears through the process of diffraction around the torso, head, and complex form of the pinnae. This diffraction can be characterized by the Head Related Impulse Response (HRIR) or equivalently by its Fourier transform, the Head Related Transfer Function (HRTF). HRTFs contain the acoustic information, such as inter-aural time differences and complex spectral cues used by the human auditory system to interpret the location of sound events. The principle is that sound arriving from any direction in space is coded by a specific pair of transfer functions (left and right ear). For a review on spatial hearing one can refer to [26]. Binaural synthesis consists in processing an audio signal by the HRTF for a given position, thus creating the virtual sound sources under headphones. Measurements of an HRTF result in a stereo filter database following a discrete spatial map. Interpolation is normally required for intermediate positions that are not in the database. More details about theses techniques can be found in [27].

Localization under static binaural rendering (no head-tracking) results in several artifacts, the most important ones being front-back confusions (a source spatialized in front of the auditor is perceived as behind, and vice versa) due to ambiguity in inter-aural differences which are symmetric relative to the inter-aural axis. The auditory system can resolve those ambiguities using head movement [16][28]. Dynamic binaural rendering, as implemented in this study, allows the exploitation of head movements (via a head mounted position tracking system) to constantly update the sound scene. One other important point is that HRTFs are dependent on human morphology and therefore an optimal binaural synthesis should be individualized to the user for better localization performances. Adaptation to non-individual acoustic cues seems to be possible [29] but requires an additional learning phase. The solution adopted in this current study was for each participant to select an "optimal" HRTF from an existing database. This procedure consisted in presenting the synthesis of a series of short sound bursts rotating about the head at a fixed distance (first in the horizontal plane, then in the median plane) using a small set of HRTFs. These HRTFs were selected from the LISTEN HRTF measurement database [30] following a perceptually significant statistical reduction procedure. The HRTF chosen by the participant as providing the most realistic source positioning, according to the known path of the sound source, was used.

A modified version of Spat was used which allowed for the individualization of inter-aural time delay, based on head circumference, independent of the selected HRTF.

5. METHOD

Fifty-four participants were selected for this study. Each one belonged to one of three groups (congenitally blind, late blind, and blindfolded sighted) and was allocated one of the two learning conditions (verbal description or active exploration). An equal distribution was achieved between the participants of the three groups according to gender, age, and educational and socio-cultural background. Each group comprised five women and four men (from 25 to 59 years of age).

5.1. Preparatory Procedure

In the preparatory phase of the experiment, which lasted approximately one hour, each participant completed a questionnaire providing information concerning their blindness origin (if any) and their socio-economic situation. Each participant then passed an audiometric exam to verify they had no appreciable hearing deficits. The head circumference of the participant was measured, to allow for individual adaptation of the ITD in the audio rendering, followed by the HRTF selection procedure.

The participants were then masked (if their vision was normal) and led into the experimental room which they were permitted to explore with the experimenter to gain an idea of its size. After this exploration, the participants were placed in front of a chair which would be a physical landmark during the experiment. The chair was located at the center of the upper wall, indicated in Figure 3 and represented by the rainbow colored sphere in Figure 1. When the participants were at this location, they knew that they were at the border of the scene, centered, that their work space was in front of them, and that the direction their body was pointed in was parallel to the lateral walls.

5.2. Learning Phase

The learning phase consisted of one of two methods for the participant to acquire the scene. Participants were moved to the center of a virtual circle (seen in Figure 3) which they were informed had a radius of 1.5 m and on which six sound sources were located. Six "domestic" sound recordings were chosen and assigned to the numbered virtual sound sources in Figure 3: (1) running water, (2) telephone ringing, (3) dripping faucet, (4) coffee machine, (5) ticking clock, (6) washing machine.

For half of the participants, the learning phase was passive and purely verbal. The participants were centered in the middle of the virtual circle and informed of the positions of the sound sources by first hearing the sound played in mono (non-spatialized), and then the experimenter verbally described its location in the terms of the conventional clock positions as used in aerial navigation, in clockwise order [14]. With only this verbal descriptive information, the participant had to form a mental representation of the whole configuration as vivid as possible.

For the second half of the participants, the learning phase was an active exploration of the spatial configuration. The participants were positioned at the center of the virtual circle. Upon the independent presentation of each sound source (correctly spatialized on the circle), they had to physically move from the center to the position of each sound source.

In order to verify that the participants correctly learned the spatial configuration, they were first placed at the center of the virtual circle. Each sound source was then played individually, non-spatialized, in random order. For the first learning condition, the participants had to express verbally where the correct source location was, in hour-coded terms. Errors were typically of the type linked to confusions between the locations of different sources rather than absolute position errors. For the second condition, the participants had to point (with the tracked pointer) to the location of the sound sources. The entire learning procedure was repeated until the responses were correct. In the pointing condition, the response was judged on the graphical display. The indicated position was valid if the pointer intersected with the sphere object. The extent of a sphere object (radius = 0.25 m) along the experimental circle

(radius = 1.5 m) equates to a range ±10° around the exact position of the sonic object.

5.3. Experimental Phase

Participants began the experimental phase at the center of the circle. They were presented briefly with one of the sound sources, non-spatialized and randomly selected, whose correct position they had to identify. To do this, the participants placed the hand-tracked pointer at the exact position in space where the sound object should be. When participants confirmed their positional choice, the sound source was activated at the position indicated and left active while subsequent sources were presented. After positioning each sound source, the participants were led back to the reference chair, and from this chair, had to position the next source. This change of reference point was intentional in order to observe the different strategies used by participants to reconstruct the initial position of sound objects. After placing the final source, all sources were active and the sound scene was complete. This was the first instance in the experiment when the participants could hear the entire scene.

Participants were then repositioned at the center of the virtual circle. They were then allowed to explore the completed scene by moving about the room. Following this, they were repositioned at the center, with the scene still active. Each sound source was selected, in random order, and participants had the possibility to correct any position they judged incorrect using the same procedure as before.

6. RESULTS

Preliminary evaluation of the experimental phase consisted in measuring the discrepancy between the original spatial configuration and the recreated sound scene. This was intended to check for any influence of the learning modality on the preservation of the metric and topological properties of the memorized environment. This discrepancy was analyzed in terms of angular, radial, and absolute distance error as compared with the correct location of the corresponding object on the periphery of the virtual circle (see Figure 3).

A detailed log file was created during the experimental phase for each participant. This log file contained the trajectory, head orientation, and other relevant actions of the participant and experimenter. Visualization of the experimental phase is thus possible using this information, of which an example is presented in Figure 6. This example is of a participant who was late blind and used the verbal learning method for acquiring the scene. The reference sound scene is shown as well as the recreated sound scene. Each action is time-stamped, as indicated by the associated index of the repositioned sources.

An analysis of errors in each learning condition will indicate whether one of the two conditions generated a more accurate spatial representation when the participants were required to report what they had learned. Any statistical dependency on blindness condition should also be apparent.

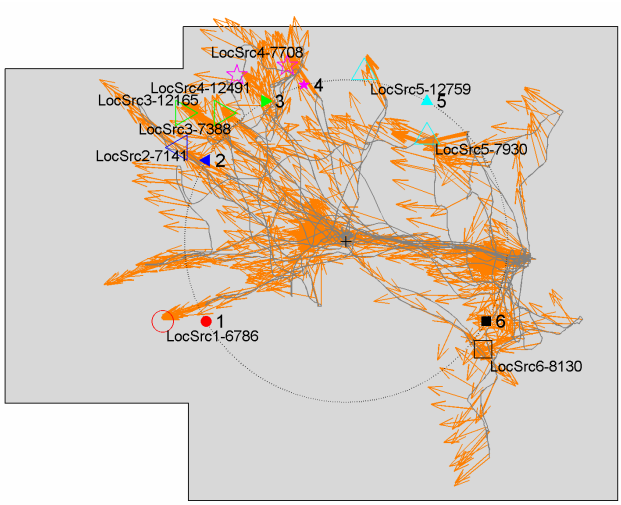

Figure 6. *Sample visualization of experimental log showing participant trajectory, head orientation (sub-sampled for clarity), and recreated scene.*

6.1. Angular Error

Angular error is defined as the absolute error in degrees, calculated from the circle center, of the position designated by participants in comparison to the reference position of the corresponding sound source. There was no significant difference between the angular error following learning from a verbal description (mean = 17°, $\sigma = 14$) and from active exploration (mean = 20°, $\sigma = 17$). Congenitally blind participants made significantly larger errors (mean = 23°, $\sigma = 17$) than late blind (mean = 16°, $\sigma = 15$) [$F(1,32) = 4.52$; $p = 0.041$] and blindfolded sighted participants (mean = 16°, $\sigma = 13$) [$F(1,32) = 6.08$; $p = 0.019$].

6.2. Radial error

Radial error is defined as the radial distance error, calculated from the circle center, between the position of the sound source and the actual position along the circle periphery. For both verbal learning and active exploration, participants underestimated the distances by the same amount (mean = 0.2 m), with similar σ (0.3 m and 0.4 m, respectively). There was no difference among the three groups; each one underestimated the distance with a mean error of 0.2 m for congenitally blind ($\sigma = 0.3$) and late blind ($\sigma = 0.4$), and a mean error of 0.1 m for blindfolded ($\sigma = 0.3$). Interestingly, a significant difference was found for blindfolded participants who underestimated radial positions when they had learned the spatial configuration from a verbal description (mean = –0.2 m, $\sigma = 0.3$) as compared with an active exploration (mean = 0.0 m, $\sigma = 0.4$) [$F(2,48) = 3.32$; $p = 0.045$].

6.3. Absolute Distance Error

Absolute distance error is defined as the distance between the original and recreated source positions. Data show a significant effect of the learning method. Active exploration of the virtual environment resulted in better absolute estimation of sound source positions (mean = 0.6 m, $\sigma = 0.3$) as compared to the verbal description method (mean = 0.7 m, $\sigma = 0.4$) [$F(1,48) = 4.29$, $p = 0.044$]. The data do not reflect any

significant difference among the groups of participants (congenitally blind, mean = 0.7 m, $\sigma = 0.4$; late blind, mean = 0.6 m, $\sigma = 0.3$; blindfolded, mean = 0.6 m, $\sigma = 0.3$).

6.4. Subjective Questionnaire

Following the experiment, all participants completed a questionnaire designed to gain insight into the performance of the system. For example, was the wall collision alert (implemented using the sound of a strong wind) efficient in preventing wall contact? The results showed that 93% responded yes, with no difference among the groups. Of specific interest was the perception of system latency, described as perceived delay between the participant's movements and the source position update, since our measure showed a significant TSL. The latency was perceived by 54% of the participants, without any difference among the groups. Systematically, in the case of perceived latency, participants stated that they adapted by reducing their speed while moving in the virtual environment.

7. DISCUSSION

The starting hypothesis was that the learning of the configuration through active exploration should better benefit blind participants than a verbal description. Such data would confirm the results of a prior set of experiments, which revealed a gain in performance of mental manipulations for blind people following this hypothesis (paper in preparation). A second hypothesis concerned sighted participants, who were expected to benefit from a verbal description since they would then be able to generate a visual mental image of the scene, and thus be able to recreate the initial configuration of the scene in a much more precise manner.

Using the task of scene recreation, the preliminary results suggest that active exploration of an environment enhances absolute positioning of sound sources as compared to learning from a verbal description. The same improvement appears with respect to radial distance errors, but only for blindfolded participants. In addition, data show that, whatever the learning modality, participants underestimate the circle width, except for the case of blindfolded, who clearly benefit from learning with perception/action coupling since the mean positioning is exactly on the 1.5 m radius circle circumference. These results are not in line with previous findings: it clearly appears that an active exploration of the environment improves blindfolded participants' performance, both in terms of absolute positioning and width of the reconstructed configuration.

We also found that blind from birth participants made significantly more angular positioning errors than late blind or blindfolded ones, and this whatever the learning modality of the environment. These data are in line with the results of previous studies involving spatial information processing in classic natural (non virtual) environments [31].

8. CONCLUSION

Contrary to previous experimental setups for audio augmented reality such as [17], the system presented in this paper offers extended facilities for the design of a reactive sonic scene. A scripting language is used to define context-dependent interactions with the experimenter and the user and complex behaviors of the sonic objects. A detailed geometrical definition of the virtual audio scene is used to present the user with a rich and easily reconfigurable sonic landscape. Last, geometrical objects can be equipped with reactive sonic properties in order to trigger warnings and avoid collisions, a necessary facility when experimenting with blind people. All these characteristics of the audio scene permitted to define an immersive and engaging virtual landscape well-suited for a detailed analysis of human auditive orientation.

The study focused on the mental imagery of representations of spatial scenes. An immersive 3D audio scene was employed, allowing for exploration of the effects of visual experience on mental processes thanks to the participation of blind and blindfolded individuals. Participants acquired knowledge of a scene either through purely verbal description or active exploration of the virtual scene. The preliminary results presented in this paper point out the interest of investigating the role of perception/action coupling in mental processing, while illustrating also how VR is a powerful tool for creating complex and interactive experimental contexts.

The interest of this study can also be viewed in relation to applications employing the presentation of spatial information through non-visual modalities. Investigating navigation with and without vision is a prerequisite to building navigation systems for blind people who are exposed to 3D audio information [32][33][34]. In this context, this work aims at complementing the now classic concept of "visual guidance of action" by reference to guiding procedures that primarily (or even exclusively) exploit audio information.

9. REFERENCES

[1] S. Lambrey and A. Berthoz, "Combination of conflicting visual and non-visual information for estimating actively performed body turns in virtual reality," *Intl. J. Psychophysiology*, vol. 50, pp. 101-115, 2003.

[2] J.M. Loomis, R.L. Klatzky, and R.G. Golledge, "Auditory distance perception in real, virtual, and mixed environments," *Mixed reality: Merging real and virtual worlds*, Y. Ohta and H. Tamura (Eds.), Tokyo, Ohmsha, pp. 201-214, 1999.

[3] I. Viaud-Delmon, A. Seguelas, E. Rio, R. Jouvent and O. Warusfel, "3-D Sound and Virtual Reality: Applications in Clinical Psychopathology," *Cybertherapy*, San Diego, 2004.

[4] I. Viaud-Delmon, L. Sarlat and O. Warusfel, "Virtual Ventriloquism: Localization of Auditory Sources in Virtual Reality," *Proc CFA/DAGA*, Strasbourg, May 2004.

[5] M. von der Heyde and H.H. Buelthoff, *Perception and action in virtual environments.* Cognitive and Computational Psychophysics Department, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany, 2000.

[6] D. Kaski, "Revision: Is visual perception a requisite for visual imagery?" *Perception,* vol. 31, pp. 717-731, 2002.

[7] N.H. Kerr, "The role of vision in "visual imagery" experiments. Evidence from the congenitally blind." *J. Exp Psycho: General*, vol.112, pp. 265-277, 1983.

[8] B. Röder and F. Rösler, "Visual input does not facilitate the scanning of spatial images," *J. Mental Imagery*, vol. 22, pp. 165-181, 1998.

[9] A. Afonso, F. Gaunet, and M. Denis, "The mental comparison of distances in a verbally described spatial layout: Effects of visual deprivation," *Imagination, Cognition and Personality*, vol. 23, pp. 173-182. 2003-2004.

[10] R. De Beni and C. Cornoldi, “Imagery limitations in totally congenitally blind subjects,” *J. Exp Psychol: Learn Mem Cogn.*, vol. 14, no. 4, pp. 650-655, 1988.
[11] A. Paivio, *Mental representations: A dual coding approach.* Oxford University Press, New York, 1986.
[12] J.T.E. Richardson, *Imagery.* Psychology Press. Hove, UK, 1999.
[13] S.M. Kosslyn, T.M. Ball, and B.J. Reiser, “Visual images preserve metric spatial information: Evidence from studies of image scanning,” *J. Exp Psychol: Human Perception and Performance*, vol. 4, pp. 47-60, 1978.
[14] M. Denis and M. Cocude, “Structural properties of visual images constructed from poorly or well-structured verbal descriptions,” *Memory and Cognition*, vol. 20, pp. 497-506, 1992.
[15] M. Denis and S.M. Kosslyn, “Scanning visual mental images: A window on the mind,” *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, vol. 18, pp. 409-465, 1999.
[16] P. Minnaar, S.K. Olesen, F. Christensen, and H. Møller, “The importance of head movements for binaural room synthesis,” *Proc ICAD*, Espoo, Finland, July 29-August 1, 2001.
[17] C. Frauenberger and M. Noisternig, “3D Audio Interface for the Blind,” *Proc ICAD*, Boston University, Boston, MA, July 7-9, 2003.
[18] VirChor. Virtual Choreographer, LIMSI-CNRS. http://virchor.sourceforge.net/
[19] MAX/MSP. Cycling’74. http://www.cycling74.com/
[20] X3D. Web3D Consortium. http://www.web3d.org/
[21] H. Hoffmann, R. Dachselt, and K. Meißner, “An Independent Declarative 3D Audio Format on the Basis of XML,” *Proc ICAD*, Boston University, Boston, MA, July 7-9, 2003.
[22] R. Dachselt and E. Rukzio, “BEHAVIOR3D: An XML-Based Framework for 3D Graphics Behavior,” *Proc ACM Web3D Symposium*, Saint Malo, France, March 2003.
[23] A. Blum, A. Afonso, B.F.G. Katz, and C. Jacquemin, “Expérimentation sur la perception de l’espace en réalité virtuelle immersive audio,” *Proc 16ème Conf Francophone sur l’Interaction Homme-Machine, AFIHM*, Namur, Belgium, August 30-Sept 3, 2004.
[24] E.M. Wenzel, “The impact of system latency on dynamic performance in virtual acoustic environments,” *Proc 16th ICA and 135th ASA*, Seattle, WA, pp. 2405-2406.
[25] E. Kahle, *Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras*. PhD thesis, Université du Maine, Le Mans, 1995.
[26] J. Blauert, *Spatial Hearing, the psychophysics of human sound localization*. MIT Press, Cambridge, 1996.
[27] D.R. Begault, *3-D Sound for Virtual Reality and Multimedia*. Academic Press, Cambridge, MA, 1994.
[28] F.L. Wightman and D.J. Kistler, “Resolution of front-back ambiguity in spatial hearing by listener and source movement,” *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2841–2853, 1999.
[29] A. Blum, B.F.G. Katz, and O. Warusfel, “Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training,” *Proc CFA/DAGA*, Strasbourg, May 2004.
[30] Listen Project – Information Society Technologies Program - IST-1999-20646: http://listen.gmd.de/ LISTEN HRTF Database: http://www.ircam.fr/equipes/salles/listen/
[31] C. Tinti, M. Adenzato, M. Tamietto and C. Cornoldi, "Visual experience is not necessary for efficient survey spatial cognition: Evidence from blindness," *Quart. J. Exp. Psychol.*, in press.
[32] R.G. Golledge, J.R. Marston, J.M. Loomis, and R.L. Klatzky, “Stated preferences for components of a Personal Guidance System for nonvisual navigation,” *J. Visual Impairment and Blindness*, vol. 98, pp. 135-147, 2004.
[33] J.M. Loomis, R.G. Golledge, and R.L. Klatzky, “Navigation system for the blind: Auditory display modes and guidance,” *Presence: Teleoperators and Virtual Environments*, vol. 7, pp. 193-203, 1998.
[34] S. Vieilledent, S.M. Kosslyn, A. Berthoz, and M.-D. Giraudo, “Does mental simulation of following a path improve navigation performance without vision?” *Cognitive Brain Research*, vol. 16, pp. 238-249, 2003.